Genome Sequencing and Analysis of a Newly Isolated Endophyte of Coffee Leaves

Research Thesis

Presented in partial fulfillment of the requirements for graduation *with research distinction* in the undergraduate colleges of The Ohio State University

By

Lauren Slattery

The Ohio State University

December 2021

Project Advisor: Dr. Jason Slot, Department of Plant Pathology

Table of Contents:

Abstract:

Endophytic fungi occupy a niche that leads to interactions with plant metabolites. In this interaction, fungi can use plant metabolites in their signaling pathways and metabolism. Fungi also must protect themselves from competing microorganisms and toxic metabolites. *Coffea arabica* was studied as a host plant due to its economic importance and diverse metabolite profile. This study identified secondary metabolite gene clusters, phenylpropanoid degrading gene clusters, and functionally annotated genes in an endophytic fungus, and highlighted genomic features that are expected in an endophytic lifestyle. Findings feature 44 metabolic gene clusters identified to be associated with secondary metabolite production and degradation. By understanding these interactions between endophytic fungi and the host plant, improvements can be made in the *Coffea arabica* cropping system in response to environmental changes and pathogens.

Introduction:

The Role of Endophytes in Tropical Plants

Endophytic fungi have a lifestyle that has unique challenges compared to soil-borne and environmental fungi as they are endosymbionts of plants. Fungal endophytes colonize plants through horizontal transmission as spores or vertical transmission as hyphae (Christian et al. 2020). In doing so, they must thrive in a chemically rich environment filled with secondary metabolites produced by the host plant in combination with their own secondary metabolites. Secondary metabolites encompass all compounds produced by organisms that are not directly connected to the growth and development of the organism. While these compounds can aid in defense and other ecological interactions, they are not strictly required for survival. The impact of endophytes in host defense is a well-understood interaction. Endophytes can trigger plant defenses and directly compete with pathogens within the host plant by producing antifungal or antibacterial compounds (Christian et al. 2020).

Environmental factors and host metabolites lead to the selection of specialized endophytes (Christian et al. 2020). Generalist endophytes are common to see across plant species, while other endophytes are adapted for host-specific preferences. Studying metabolites can increase the understanding of endophytes and plants in a variety of conditions from pathogenic to beneficial (Christian et al. 2020).

Given the biodiversity of tropical ecosystems, plants endemic to tropical regions have a high potential for producing undiscovered endophyte secondary metabolites (Higginbotham et al. 2013). In a study by Higginbothan et al. (2013), fungal endophytes from various lineages living in association with leaves were isolated from angiosperms native to Panama. Extracts of culture

supernatants were tested for their bioactivity in vitro against four causes of human disease: *Plasmodium falciparum* (causal agent of malaria), *Leishmania donovani* (causative agent of leishmaniasis), *Trypanosoma cruzi* (causal agent of Chagas disease), and human breast cancer cell line MCF-7. Seventeen percent of the isolates per assay were rated as highly active against the causal agent studied, underscoring the potential of plant endophyte metabolites in treating human disease. The compounds present in each isolate with high activity are still being identified (Higginbotham et al. 2013).

The Diverse Metabolites of the *Rubiaceae* Family

Rubiaceae is a highly diverse family of tropical angiosperms. This family contains species that produce important chemical substances, including *Coffea arabica* (caffeine), *Psychotria viridis* (dimethyltryptamine), and *Mitragyna speciosa* (mitragynine). Indigenous cultures have also been using secondary metabolites from the *Rubiaceae* family to treat diseases for generations (Cruz et al. 2020). A particularly important indigenous medicinal plant is *Cinchona officinalis* which produces the antimalarial quinine. Given this history, the diversity of secondary metabolites found in *Rubiaceae* is a continuing field of study. In modern medicine, *Uncaria tomentosa* has been studied for its production of metabolites with anticancer properties. In one study, alkaloid extracts from the bark of *U. tomentosa* were successful in inducing apoptosis in promyelocytic leukemia and thyroid carcinoma (García & Grijalva 2021).

Endophytic fungi found in *Rubicaeae* have been studied for potential medical uses of their secondary metabolites. For example, a study from Cruz et al. (2020) believed to have isolated the anticancer drug, Taxol, as a secondary metabolite from the fungus *Botryodiplodia theobromae* isolated from *Morinda citrifolia*, and *Aspergillus oryzae* isolated from *Tarenna asiatica*. However, further studies should be done as the hypothesis of fungal production of

Taxol is controversial. Also, the antibiotics brefeldin A and trichodermol were isolated from a fungal endophyte of *Scyphiphora hydrophyllaceae*. Secondary metabolism of endophytic fungi of *Rubiaceae* is an expanding field of study as there are emerging opportunity to isolate potentially medically useful secondary metabolites from *Rubiaceae* endophytes (Cruz et al. 2020).

Lasiosphaeriaceae: An Emerging Fungal Family of Study

Lasiosphaeriaceae is a large, diverse paraphyletic family in the order Sordariales (Marin-Felix et al. 2020). Many fungi in this family inhabit plant debris and soil (Marin-Felix et al. 2020). Lasiosphaeriaceae species have also been found as plant endophytes. Specifically, *Cercophora samala* was found to be an endophyte of *Mitragyna inermis* (Rubiaceae). Two new secondary metabolites, mitrafungidione and maristachone F were identified in this fungus (Vouffo et al. 2021). Coprophilous (dung inhabiting) fungi from Lasiosphaeriaceae have also been studied for their secondary metabolites including antifungal compounds (Sarrocco 2016). Decipin A, an antifungal compound produced by *Podospora decipiens,* was found to restrict growth of the plant pathogen, *Fusarium verticillioides*. Important for human health, *Cercophora areolata* was found to produce Isocoumarin decarboxy-citrinone, which inhibits the growth of *Candida albicans,* a human pathogen (Sarrocco 2016).

Plant Secondary Metabolite Production in Endophytic Fungi

Some plant metabolites are thought to also be produced by their associated endophytes (Sachin et al. 2013). For example, *Diaporthe* sp., an endophyte of *Cinchona ledgeriana*, was identified to produce quinine in culture (Maehara et al. 2012). In another study, there was genetic evidence of plant secondary metabolite genes in fungi (Salazar-Cerezo et al. 2018). However,

when grown in culture, the fungus did not produce the identified metabolites. This is believed to be caused by lack of host stimulus. Terpenoid indole alkaloids (TIAs) are a large group of plant secondary metabolites found in *Apocynaceae*, *Loganiaceae*, and *Rubiaceae*. TIA production in endophytes may be due to convergent evolution as the production of plant hormones may aid in pathogenesis of fungi (Salazar-Cerezo et al. 2018).

Metabolic Gene Clustering in Fungi as an Indicator of Niche

Metabolic gene clusters (MGCs) are co-localized genes that participate in the same metabolic pathway (Wisecaver & Rokas 2015). They can take the form of biosynthetic pathways, degradation pathways, or others (Slot 2017). MGCs increase the ability of fungi to metabolize nutrients using secreted and transporter proteins as well as produce antimicrobial compounds. MGCs can be similar in distantly related species that share niches, suggesting that MGCs are affected by ecological selection (Gluck-Thaler & Slot 2018). The presence of MGCs in distantly related organisms is due to horizonal gene transfer (HGT). HGT is the transfer of genes between organisms by a non-reproductive process. The exact mechanisms of genetic material uptake and incorporation in HGT by an organism is yet to be characterized. In addition, MGCs in fungi producing secondary metabolites are diverse. However, specific clusters tend to have narrow taxonomic distributions (Wisecaver & Rokas 2015). By studying MGCs, the potential interactions of a fungus in its environment may be better understood.

The Role of Endophytic Secondary Metabolites in Plant Health Management

Endophytic fungi have been found to produce metabolites that help host plants tolerate biotic and abiotic stress (Monteiro et al. 2016). In some instances, endophytic fungi can reduce the growth of pathogenic fungi. In the case of *Coffea arabica*, the endophytic fungus *Muscodor albus*

was found to produce volatile organic compounds that reduced the growth of 378 of 400 tested fungi, including the plant pathogen *Fusarium verticilliodes* (Monteiro et al. 2016). Endophytes may be a source of biocontrol agents or compounds for promoting plant health. Understanding interactions between endophytes and *Coffea arabica* may play a role in supporting the health of the internationally valuable coffee industry.

This study aimed to find secondary metabolite production and degradation gene clusters as well as other ecologically important genes in the genome of a Lasiosphaeriaceae endophyte isolated from *Coffea arabica.* This study identified 39 putative fungal secondary metabolite gene clusters. It also identified phenolic catabolism clusters including quinic acid cluster and pterocarpan hydroxylase cluster. Potentially ecologically important genes associated with the fungus' endophytic lifestyle included an auxin mediated pathway, purine transporter, deaminase, molybedenum cofactor sulfurtransferase, and multiple cutinase genes.

Materials & Methods:

Culturing & ITS Identification

Two young coffee plants were acquired from a Florida Nursery (Plantvine.com). Upon receipt, their identity was confirmed by sequencing the ITS region using ITS1 & ITS4 primers according to protocol (Grilanda et al. 2011). The Sanger sequencing results were compared against known sequences using NCBI nucleotide Basic Local Alignment Tool (BLAST). The resulting sequence from the database that had the highest percent identity was from *Coffea arabica*. The coffee plants were maintained in the laboratory for four weeks, watered with sterile 90% milliQ: 10% tap water weekly, prior to endophyte isolation.

The endophyte isolation protocol is as follows: one young leaf was selected from each plant from the second set of leaves down from the apical meristem. The leaves were surface sterilized by sequentially submerging the leaves in 3.5% bleach for 30 seconds, 70% ethanol for 30 seconds, and three rinses of autoclaved miliQ water for 30 seconds each. From the surface sterilized leaves, 10 mm sections of leaf were excised using a sterile scalpel and placed on potato dextrose agar (PDA) media in a petri dish (Figure 1). A sample of the final sterile water wash was added to a PDA plate to serve as a negative control for any contamination. The cultures were left to grow in a 25°C incubator. They were monitored for fungi of different morphology. Any fungal growth with visible differences in morphology were separated by subculturing onto new PDA plates. Once a 1/2" diameter circle of mycelium formed a pure culture, containing only one fungus species as recognized by visual morphology, the fungus was identified by sequencing its ITS region.

Endophytic fungi were identified by sequencing of the ITS1F and ITS4 primer pair amplicon. The reaction included a denaturing step of 93°C for 4 minutes followed by 35 cycles of 94°C for 45 seconds, 57° for 45 seconds and 72°F for 1 minute. The final extension was 72°F for 7 minutes. The resulting PCR product was sequenced using Sanger sequencing (Ohio State University Genomics Shared Resource) and compared to the NCBI database using nucleotide BLAST. The sequences with the highest percent identity to the unknown endophytes are shown in table 1.

DNA Extraction & Sequencing

Culture CE_1 identified as *Cercophora* sp. was selected for further study because it was repeatedly isolated and the genus is not well studied. A block of agar was transferred to potato dextrose broth (PDB) in a sterile flask and allowed to grow for 14 days on a shaker, at room temperature in the dark. The fungal tissue was removed from the broth, rinsed with two consecutive rinses of 50 ml of sterile miliQ water and blotted dry with a sterile paper towel. The tissue was frozen using liquid nitrogen and macerated using a sterile mortar and pestle. DNA extraction of the fungal tissue was performed using the Qiagen Plant Mini Kit (Hilden, Germany), with the standard protocol. The extracted DNA was tested for quality and concentration using Nanodrop and Qubit. A sample of 29.90 ng/μL of DNA was sent for Illumina sequencing at Novogene using NovaSeq 6000 with the PE150 strategy.

Genome Assembly

Downstream analysis of the whole-genome sequencing results was completed using The Ohio Supercomputing Center (www.osc.edu). The raw sequencing data from Novogene was assessed using FastQC (v0.11.8 Andrews et al. 2012) and Trimmomatic (v0.3.6 Bolger 2014)

was used to remove adapter content found at the ends of the reads. FastQC was used again to ensure the Trimmomatic step was effective in removing adaptors. Next, the genome was assembled using SPAdes genome assembler (v3.14.1, Prjibelski et al. 2020). A QUAST (v4.6.3 Gurevich et al. 2013) report was generated to visualize the assembly statistics to ensure the assembly was successful. The assembly statistics are summarized in table 4.

Genome Annotation & Analysis

To begin the genome annotation, RepeatModeler (v1.0.11 Flynn et al. 2020) was used to generate a library to recognize repetitive elements in the assembled genome referencing the Dfam repetitive element database (v3.0 Wheeler et al. 2012). The Funannotate pipeline (v1.7.4 Palmer 2015) was used in conjunction with SNAP (v2006-07-28 Korf 2004), AUGUSTUS (v3.3.3 Stanke et al 2004), GlimmerHMM (v.3.0.4 Majoros 2004), and GeneMark-ES (v4.35 Lukashin & Borodovsky 1998) to predict genes from the assembly. First, the assembly was sorted to remove contigs below 1kB and sort them from largest to smallest. The genome was soft masked using a library produced from *Lasiosphaeriaceae* family fungi to detect repeating elements in the genome. Next, a database was generated to train the Funannotate gene prediction software using benchmarking universal single-copy orthologs (BUSCO) (v3 Simão et al. 2015). Then, Funannotate used the masked assembly, BUSCO database, and protein evidence from ten closely related species within *Lasiosphaeriaceae* to identify proteins from the assembled genome and generate an annotated genome (table 2). Most were found inhabiting either dung or dying wood.

To improve the annotation quality, OrthoFiller (v1.1.4 Dunne & Kelley 2017) was used to refine gene identification. The OrthoFiller output was analyzed using antiSMASH (v6.0.1 Blin et al. 2021) to identify secondary metabolite gene clusters. Putative functions were assigned to

the Orthofiller output using Eggnog. (v5.0 Huerta-Cepas et al. 2019). The functional annotations were searched for keywords associated with an endophytic lifestyle including: auxin, gibberellin, cytokinin, salicylic acid, ethylene, jasmonate, brassinosteroid, caffeine, xanthine, cutin, cuticle, cellulose, pectin, and lignin. Cluster retrieve was used to recognize phenylpropanoid degrading clusters (Gluck-Thaler & Slot 2018) using scripts found at github.com/egluckthaler/cluster_retrieve. Lastly, a phylogenetic tree was generated based on RNA polymerase II (Rpb2) genes of closely related species in *Lasiosphaeriaceae* (table 3) IQ tree (v1.6.11 Nguyen et al. 2015). The most closely related genome to CE_1 was *Lasiosphaeriaceae* sp. with a bootstrap value of 61. *Lasiosphaeraceae sp.* and CE_1 joined a clade with *Schizothecium vesticola, Cercophora newfieldinana, Cercophora caudata,* and *Podospora curvicola* with a bootstrap value of 100. The produced tree is not included as most of the genomes used were unpublished.

Results:

Three different genera of fungi were isolated from leaves of *Coffea arabica*: *Chaetomium* (LC639849.1), *Cercophora* (MW114387.1), and *Sordiaromycetes* (JQ760297.1). Figure 1 shows how pieces of surface-sterilized *Coffea arabica* leaves were plated on PDA, where fungus can be seen growing from the plant tissue. Figure 1 A & B show a pure culture of isolate CE_1 that was used for ITS identification and making backup cultures. Of the three isolates found in table 1, CE_1 was chosen for full genome sequencing as it was repeatedly isolated. CE_1 had the highest nucleotide identity (99.8%) to an unknown endophyte isolated from *Olyra latifolia*. The highest nucleotide identity (97.47%) to an isolate of known identity was to *Cercophora fici*, isolated from *Ficus ampelos* leaf litter. (Tennakoon et al. 2021).

The assembled genome was 41,770,949 bp contained in 181 contigs (table 4). Following annotation with Funannotate, 13,435 genes were identified. The use of Orthofiller increased the number of potential genes by 2.1% as 283 additional genes were identified. 44 potential gene clusters were identified. Of these, 39 were secondary metabolite gene clusters identified by antiSMASH (figure 2). Most secondary metabolite gene clusters were identified as Type 1 polyketide synthase (T1PKS). There was also a variety of non-ribosomal peptides (NRPS), terpene, ribosomally synthesized and post-translationally modified peptides (fungal-RIPP), indole, and combination clusters. Cluster retrieve scripts were used in identifying five putative catabolic gene clusters (figures 6,7,13,14, & 15) (Gluck-Thaler & Slot 2018). A quinic acid degradation cluster was identified which is known to be highly conserved in fungi (figure 7). A pterocarpan hydroxylase cluster was identified which may be involved in the breakdown of host plant defenses (figure 8). Three additional clusters of interest were identified for potential catabolism of plant metabolites (appendix B).

Five genes potentially related to an endophytic lifestyle were identified (figures 3-5) from the functional annotation using Eggnog based on keywords related to known plant compounds and metabolites stated in the methods. An auxin-activated signaling pathway gene (figure 3) was found near a sodium-hydrogen exchange factor which is used in pH regulation (Dragwidge et al. 2018).

Five putative cutinase genes were also found in the genome. One cutinase gene was found downstream of a purine transporter and deaminase (figure 5). Four other cutinase genes were found on different contigs (appendix a) molybdenum cofactor sulfurase gene (figure 6) that could support the function of xanthine dehydrogenase was identified in two different locations. However, no xanthine dehydrogenase gene was found using the keywords caffeine, xanthine, purine, deaminase, or alkaloid.

Discussion:

Ecological Selection Through Metabolic Gene Clustering

To investigate genetic characteristics that are indicative of an endophytic lifestyle, metabolic gene clusters (MGCs) were studied. Pathways used in environmental interactions by fungi may be better understood by identifying metabolic gene clusters. Gene clusters are physically linked genes that contribute to a related function (Wisecaver, Slot, & Rokas 2014). While there are metabolic gene clusters involved in vital growth functions, others produce and degrade specialized (or secondary) compounds to aid in environmental interactions (Slot 2017).

Endophytic fungi exist on a spectrum from beneficial to pathogenic to their host organism based on the host's identity and environmental conditions, leading to condition-specific secondary metabolite production (Christian et al. 2020). Ecological selection has led to highly regulated gene clusters for secondary metabolite pathways. In an endophytic lifestyle, the expression of enzymes that break down plant secondary metabolites is necessary for survival. Secondary metabolism in fungi can inhibit the growth of competitors and be used in communication by chemical signals. (Slot 2017). Therefore, MGCs are a useful component of an endophytic lifestyle.

Potential Antimicrobial Properties of Produced Secondary Metabolites

Polyketides are a group of natural products produced by a variety of organisms by polyketide synthases (Abdalla et al. 2020). They have been studied for antimicrobial, anticancer, immune-suppressing, anti-viral, anti-inflammatory, and anti-cholesterol properties. Under appropriate conditions, endophyte produced polyketides may contribute to host plant antimicrobial activity. The most common type of secondary metabolite gene cluster identified in CE_1 was type

I polyketide synthase (T1PKS) (figure 2). This is consistent with another Rubiaceae endophyte as *Talaromyces funiculosus*, isolated from *Psychotria zombamotana* leaf tissue that was identified to have multiple (T1PKS) gene clusters (Abdalla et al. 2020).

Non-ribosomal peptides (NRP)s are produced by specialized enzymes and do not require ribosomes (Deepika et al. 2015). Some NRPs are used for vegetative growth and reproduction while others interact with host organisms or the environment (Oide & Turgeon 2020). In addition, seven terpene, one indole, and one fungal-RIPP gene clusters were identified in addition to combination clusters. None of the secondary metabolite gene clusters had close matches to known clusters to further identify their function. As an endophyte, these secondary metabolite gene clusters may be useful in interfacing with the host plant.

Putative Plant Secondary Metabolite Degrading Genes

In an endophytic lifestyle, fungi likely utilize plant secondary metabolites as carbon sources and for their own metabolism. Phenylpropanoid degrading clusters, a type of secondary metabolism gene cluster, were identified (Figure 7, 8, & appendix B). MGCs have also been studied for their role in degrading plant secondary metabolites (Gluck-Thaler & Slot 2018). As one of the largest groups of plant secondary metabolites is phenylpropanoids, this has been the subject of searches for specific MGCs (Gluck-Thaler & Slot 2018). Phenylpropanoid degrading pathways were inferred by testing for unexpectedly conserved linkage to functionally characterized "anchor genes". Anchor genes act as a starting location where gene clusters surrounding that gene are checked for similar functions (Gluck-Thaler & Slot 2018).

The presence of multiple types of degradation clusters may be indications as to the lifestyle of the isolated fungus. Five phenylpropanoid degrading clusters were identified in CE_1. A quinic acid catabolism cluster which is highly conserved in fungi as a way of obtaining

carbon was identified (Figure 7) (Valone, Case, & Giles 1971). A pterocarpan hydrolase cluster (Figure 8) may be used in the degradation of pterocarpan and medicarpin isoflavonoids (Durango et al. 2014). Isoflavonoids have been identified for defense against fungi in Fabaceae. This gene cluster also has two cytochrome P450 genes which may be used in the detoxification of pterocarpan and medicarpin isoflavonoids (Gluck-Thaler & Slot 2018). The products of these genes could allow fungi to degrade toxic compounds that are part of the host plant's defense. The presence of three additional phenylpropanoid degrading clusters demonstrate that the isolate may encounter a variety of secondary metabolites that it can utilize from its environment or needs to degrade for its survival (appendix B).

Caffeine is a xanthine alkaloid produced by *Coffea arabica* with known antifungal properties, so it might be expected for a coffee endophyte to contain enzymes to degrade or detoxify caffeine. From the functional annotations, a purine transporter and deaminase were located downstream of a cutinase gene (figure 5). The purine transporter and deaminase genes may be used in a pathway in conjunction with the molybdenum cofactor sulfurtransferase genes (figure 6). While not identified in CE_1, xanthine dehydrogenase requires a molybdenum cofactor to function in the oxidative metabolism of purines. Following purine catabolism, end products become carbon and nitrogen sources for the organism (Dwivedi et al. 2013). This indicates the potential presence of genes necessary for the breakdown of caffeine. The role of caffeine in plant biology takes two forms: allelopathy and priming. Allelopathy is the secretion of chemicals from a plant to inhibit the growth of surrounding organisms. Caffeine also affects signaling pathways, which indirectly stimulates plant defenses (Sugiyama et al. 2016). Therefore, endophytic fungi may benefit from degrading antifungal caffeine using the mentioned pathway involving xanthine dehydrogenase (Schulz et al. 2013).

Niche Indications by Reported Auxin Mediated Pathway and Cutinase Genes

In an endophytic lifestyle, auxin synthesis, transport, and signaling have been found to be components of plant-microbial interactions especially when the endophyte lives in association with roots (Sukumar et al. 2013). The exact use of auxin in fungi is yet to be understood, but auxin mediation genes have been identified. CE_1 contains a putative auxin mediated pathway gene which may indicate that the isolate responds to auxin (figure 3). Downstream from this gene is a sodium hydrogen exchanger which is important in pH homeostasis. While these two genes are not a cluster in plants, they are known to cause pH changes in plants which causes changes in auxin gradients (Dragwidge et al. 2018). Both the sodium hydrogen exchanger and auxin mediated pathway genes could be useful in an endophytic lifestyle.

Cutinases are enzymes that digest cutin, a substance forming the surface layer of plants (Yao & Köller, 1993). Studies in epiphytes and endophytes have determined that the presence of multiple cutinase genes shows an adaptation for the plant surface niche. Multiple cutinase gene presence is also common in pathogenic fungi. For example, *Magnaporthe oryzae* had 18 putative cutinases in one study (Wang et al. 2017). Four potential cutinases genes were identified in the genome of CE_1. In one copy of the gene, a hydrolase was located upstream which has been identified in cutinase gene clusters before (Wang et al. 2017). This further justifies the potential for CE_1 to be an endophyte due to the presence of multiple genes associated with known plant-inhabiting fungi.

Conclusion:

Clues indicating an endophytic lifestyle were identified by investigating the genome of a fungal isolate discovered in coffee leaves. The approaches included: predicting secondary metabolite genes, reviewing functional gene annotations, and searching for catabolic gene clusters. When a fungal endophyte, isolated from *Coffea arabica*, was sequenced, 13,718 genes were identified. There were 39 potential gene clusters identified to be involved in secondary metabolite production by the isolate. Five major phenylpropanoid degrading clusters were identified which connect to fungal breakdown of plant metabolites. Notably, the quinic acid degradation cluster was identified as expected as it is highly conserved in fungi and the pterocarpen hydroxylase which may be useful in breaking down host defense to fungi. From the functional annotations, five different genes were studied as they may contribute to an endophytic lifestyle. The auxin signaling pathway gene indicates that the isolate may be interacting with auxin levels within the plant. This could be used to determine the localization or growth of the endophyte. The cutin hydrolysis gene could provide a method of fungal introduction to the host plant through the breakdown of the plant cuticle. Purine transport and deaminase genes could be used by the endophyte in the breakdown of xanthine alkaloids produced by the plant, such as the antifungal metabolite, caffeine. Studying the biochemical interactions occurring between *Coffea arabica* and an endophytic fungus inhabiting it through the analysis of the endophyte's genome, may lead to improved methods of coffee plant management and reveal new metabolites to be used in advancing human health.

Tables:

Table 1: Identified Endophytes from Coffee Plant

Isolated Endophyte	Top BLAST Hit Name	Accession Number	Percent Identity	Top Identified BLAST Hit Name	Accession Number	Percent Identity
CE_1	Fungal endophyte isolate	EU686840.1	99.8	Cercophora fici	MW114387.1	97.47
CE_2	Fungal endophyte isolate	KY022839.1	99.31	Chaetomium globosum	LC639849.1	99.31
CE_3	Sordariomycetes sp.	JQ760297.1	100	Apodus oryzae	MK120120.1	99.8

Table 2: JGI Reference Database Used by Busco

Name	Strain Number	Sample Site
Apiosordaria backusii	CBS540.89	rabbit dung
Apodospora peruviana	CBS118394	wombat dung
Bombardia bombarda	SMH3391-2	log, cut end & branches
Cercophora caudata	CBS606.72	soil
Cercophora newfieldiana	SMH2532-1	wood fragment
Cercophora scortea	SMH4131-1	hardwood branch
Lasiosphaeris hispida	CBS955.72	decaying wood
Lasiosphaeria miniovina	SMH2392-1A	tree branch
Lasiosphaeria ovina	CBS958.72	*Fagus sylvatica* decaying wood
Schizothecium vesticola	SMH3187-1	deer dung

Table 3: Isolates used in making an Rpb2 Tree Sourced from JGI

Name	Strain number
Apiosordaria backusii	CBS 540.89
Apodospora peruviana	CBS 118394
Bombardia bombarda	SMH3391-2
Cercophora caudata	CBS 606.72
Cercophora newfieldiana	SMH 2532-1
Cercophora scortea	SMH 4131-1
Lasiosphaeria ovina	CBS 958.72
Lasiosphaeria miniovina	SMH 2391-1A
Lasiosphaeriaceae sp.	AZ0830
Lasiosphaeris hisuta	SMH4607-1
Lasiosphaeris hispida	CBS 955.72
Podospora appendiculata	CBS 314.62
Podospora curvicolla	TEP21a
Podospora didyma	CBS 232.78
Schizothecium vesticola	SMH 3187-1

Table 4: Annotation & Assembly Statistics for CE_1

Assembly Statistics

N50	816229
L50	15
L50%	0.08287
Largest Contig	1975155
Shortest Contig	1000
Contigs	181
Assembly Length	41770949
GC	0.5479
Mask %	3.0269

Funannotate Annotation Statistics

Gene Length	21196361
Number of Genes	13435
Mean Gene Length	1575.586
Median Gene Length	1312

Orthofiller Annotation Statistics

Gene Length	21470230
Number of Genes	13718
Mean Gene Length	1565.114
Median Gene Length	1302

Figures:

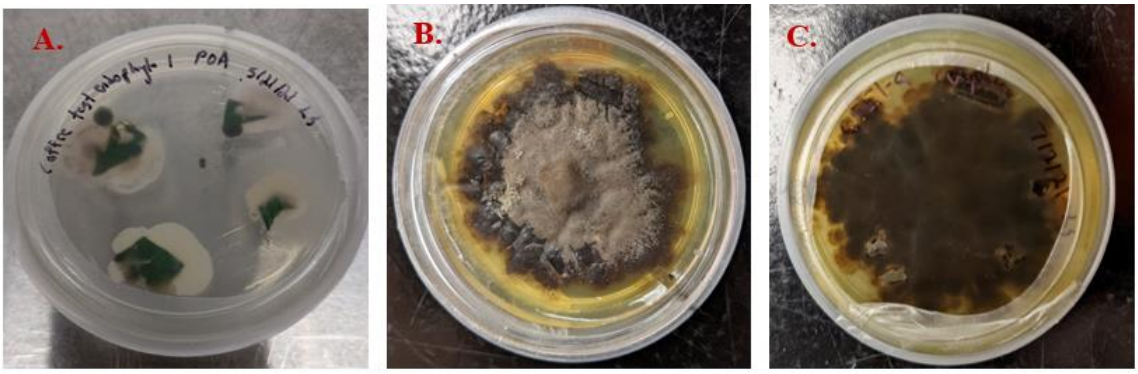

Figure 1: (A) Surface Sterilized Coffee Leaves on PDA. There is no visible contamination at this point as all fungal growth is originating at the four leaves. (B & C) Top and bottom of one plate that was categorized as a pure culture of CE_1 on PDA. This plate was used in ITS identification and making of flasks of tissue grown in PDB for DNA sequencing.

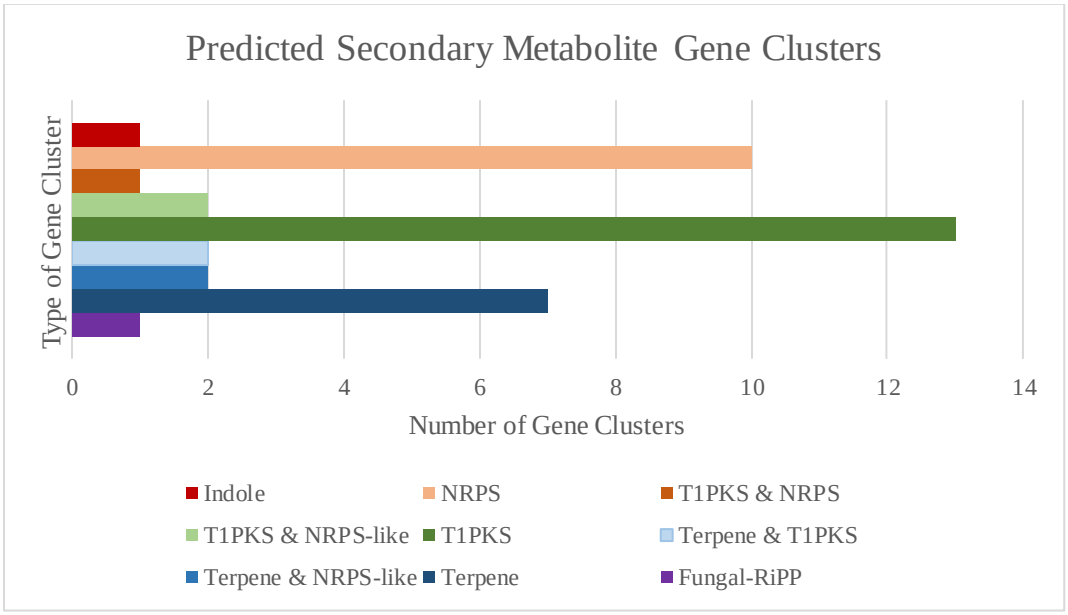

Figure 2: The results from AntiSMASH predicted secondary metabolite gene clusters are shown. A total of 39 potential secondary metabolite gene clusters were identified in CE_1. There was a combination of Indole, NRPS, T1PKS, Terpene, and Fungal-RiPP clusters. None closely matched known gene clusters for further identification.

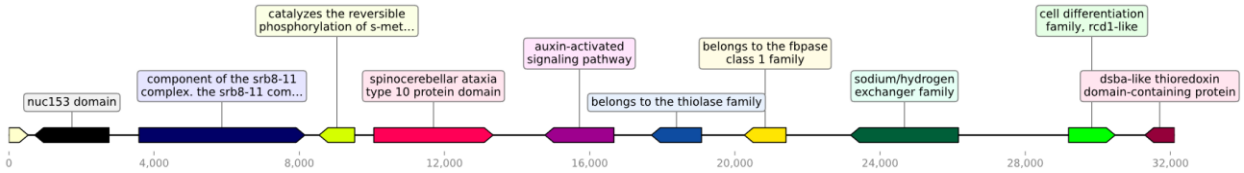

Figure 3: The auxin-activated signaling pathway in the figure was identified from functional annotations by Eggnog. It is upstream of a sodium & hydrogen exchanger family. The products of these genes are known to interact in hormone pathways in plants. The link between these proteins in fungi is unknown.

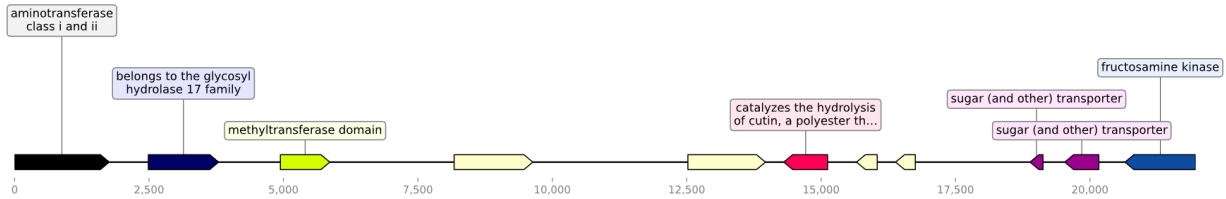

Figure 4: A hydrolase (blue) is located upstream of a cutinase gene (red) were identified from functional annotations by Eggnog. These two genes may interact in the breakdown of plant cuticle.

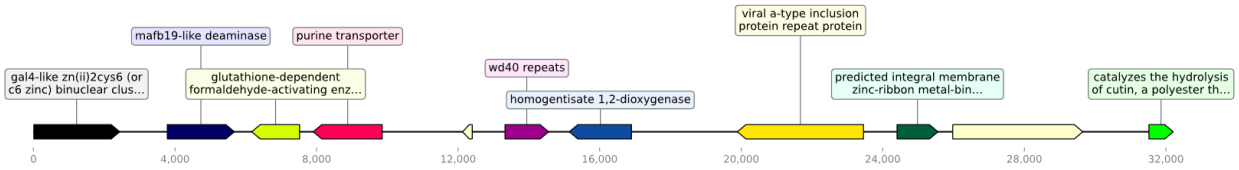

Figure 5: A deaminase (blue), purine transporter (red), and cutinase (green) were identified from functional annotations by Eggnog. The deaminase and purine transporter may interact in the same xanthine breakdown pathway in fungi. The cutinase gene indicates that CE_1 may encounter cutin in its environment.

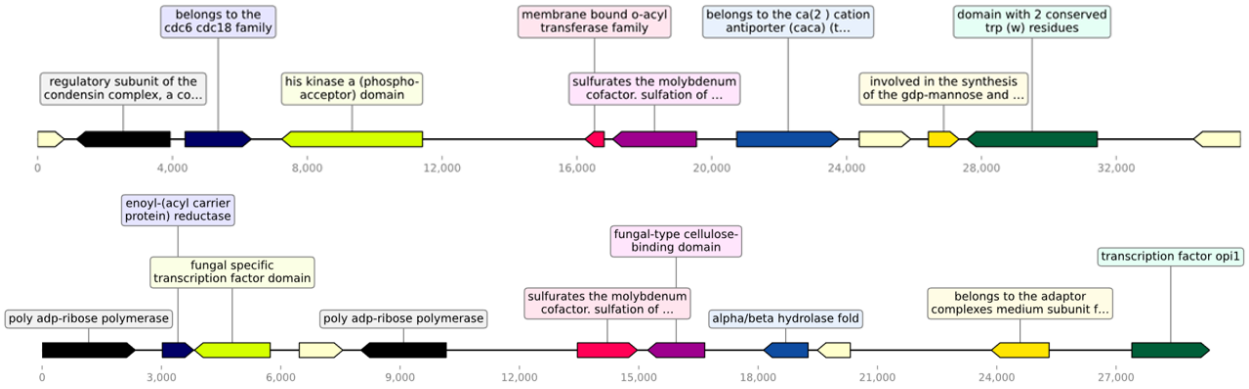

Figure 6: Two molybdenum cofactor sulfurtransferase loci (purple in top and red in bottom) were identified from functional annotations by Eggnog. These enzymes may function in the same pathway as xanthine dehydrogenase. However, the xanthine dehydrogenase gene was not identified in CE_1.

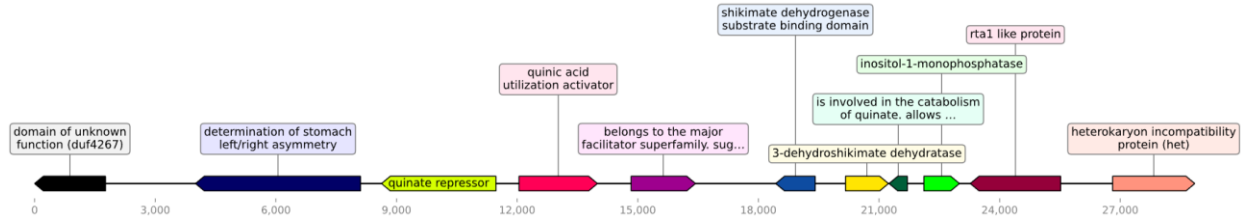

Figure 7: This gene cluster is used in quinic acid degradation and was identified using cluster retrieve scripts. This cluster is highly conserved in fungi.

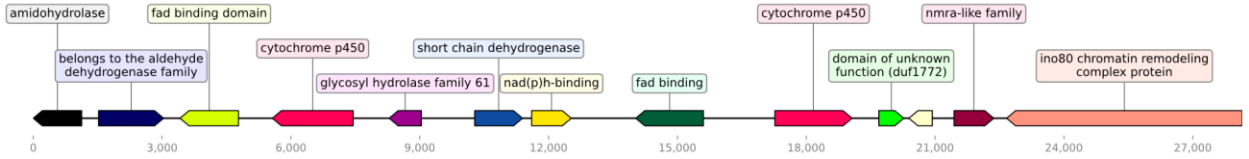

Figure 8: The pterocarpan hydroxylase gene cluster was identified using cluster retrieve scripts. It is thought to be used by fungi to break down host plant defenses for endophyte survival or pathogenicity.

Appendix A:

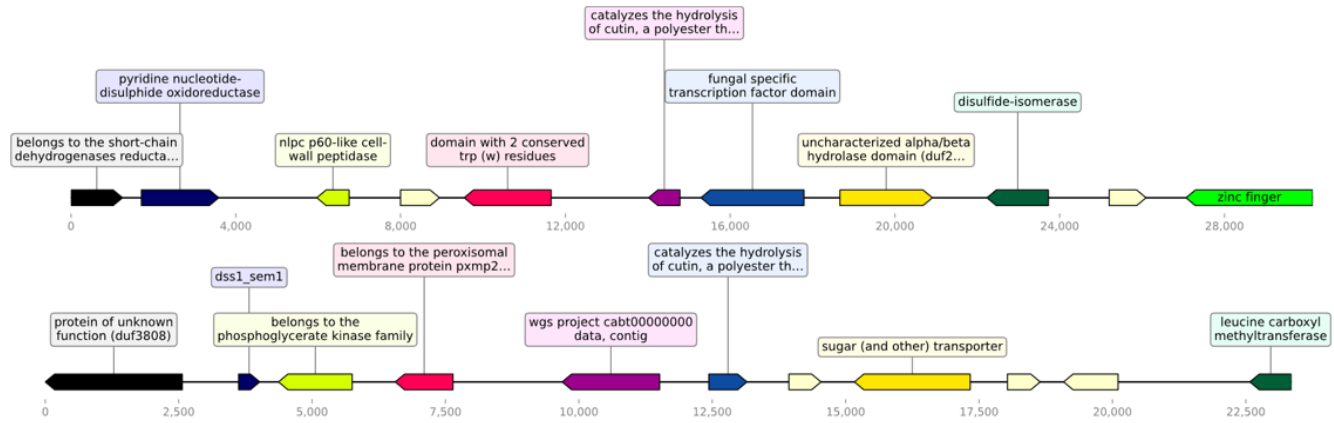

Figure 9: Additional cutinase gene loci were identified from functional annotations by Eggnog. They were found on various contigs, and each have different surrounding genes.

Appendix B:

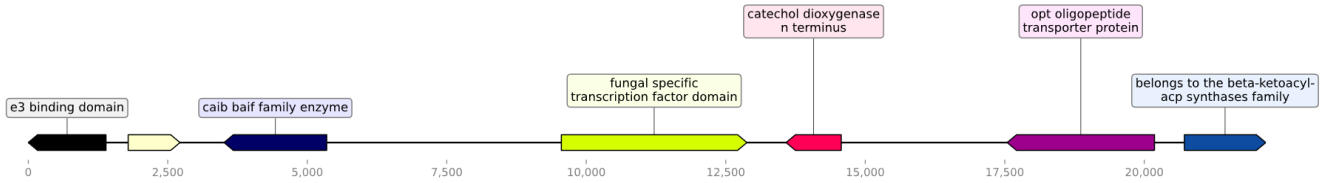

Figure 10: The catechol dioxygenase gene cluster was identified using cluster retrieve scripts.

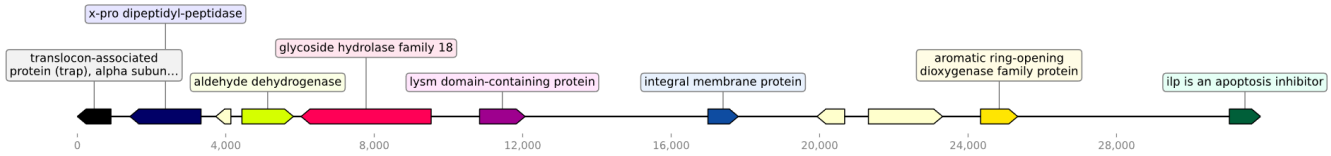

Figure 11: The aromatic ring-opening dioxygenase gene cluster was identified using cluster retrieve scripts.

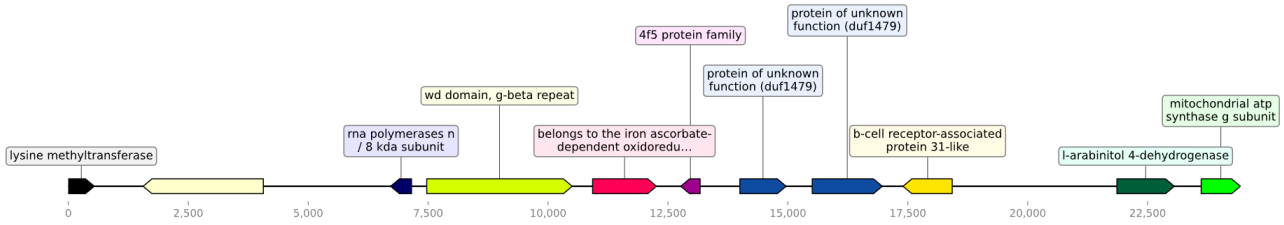

Figure 12: The naringenin 3-dioxygenase gene cluster was identified using cluster retrieve scripts.

Acknowledgements:

Foremost, I am thankful to Kelsey Scott for her mentorship over the past three years. I am grateful to Dr. Jason Slot for the opportunity to grow my skills as a mycologist in his lab. I am appreciative of Zachary Konkel for his time in teaching me how to use the genome annotation software. I appreciate the input of Matthew Meyer and Guillermo Valero David in revisions. Lastly, I would like to thank Ian Parry, Emma Slattery, Tracey Slattery, and Craig Slattery for their moral support.

References:

Abdalla, M. A., Aro, A. O., Gado, D., Passari, A. K., Mishra, V. K., Singh, B. P., & McGaw, L. J. (2020). Isolation of endophytic fungi from South African plants, and screening for their antimicrobial and extracellular enzymatic activities and presence of type I polyketide synthases. South African Journal of Botany, 134, 336-342. doi.org/10.1016/j.sajb.2020.03.021

Andews, S., Kruger, F., Segonds-Pichon, A., Biggins, L. Krueger, C., & Wingett, S. (2012). FastQC. *Babraham Institute*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., & Tilmann, W. (2021). antiSMASH 6.0. *Nucleic Acids Research*. doi:10.1093/nar/gkab335

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi.org/10.1093/bioinformatics/btu170

Christian, N. Sedio, B. E., Florez-Buitrago, X., Ramirez-Camejo, L. A., Rojas, E. I., Mejia, L. C., Palmedo, S., Rose, A., Schroeder, J. W., & Herre, E. A. (2020). Host affinity od endophytic fungi and the potential for reciprocal interactions involving host secondary chemistry. *American Journal of Botany*. 107.(2), 219-228. doi:10.1002/ajb2.1436

Cruz, J. S., Amaral da Silva C., & Hamerski, L. (2020). Natural Products from Endophytic Fungi Associated with Rubiaceae Species. *Journal of Fungi*. 128.(6), doi:10.3390/jof6030128

Deepika, V. B., Murali T. S., & Satyamoorthy, K. (2016). Modulation of genetic clusters for synthesis of bioactive molecules in fungal endophytes: A review. *Microbiological Research*. 182, 125-140. doi.org/10.1016/j.micres.2015.10.009

Dragwidge, J. M., Ford, B. A., Ashnest, J. R., Das, P., & Gendall, A. R. (2018). Two endosomal NHX-type Na+/H+ antiporters are involved in auxin-mediated development in Arabidopsis thaliana. *Plant and Cell Physiology*, 59(8), 1660-1669. doi.org/10.1093/pcp/pcy090

Durango, D., Pulgarin, N., Jesús, G. I. L., Escobar, G., Echeverri, F., & Quiñones, W. (2014). Differential accumulation of defense-related isoflavonoids in hypocotyls/roots of common bean (Phaseolus vulgaris L.) cultivars treated with salicylic acid and structurally related compounds. *Boletín Latinoamericano y del Caribe de Plantas Medicinales y Aromáticas*, 13(4), 381-405. ISSN 0717 7917

Dunne, M. P., & Kelly, S. (2017). OrthoFiller: utilizing data from multiple species to improve the completeness of genome annotations. *BMC genomics*, 18(1), 1-16. doi.org/10.1186/s12864-017-3771-x

Dwivedi, V. D., Arora, S., Kumar, A., & Mishra, S. K. (2013). Computational analysis of xanthine dehydrogenase enzyme from different source organisms. *Networking Modeling Analysis in Health Informatics and Bioinformatics*, 2, 185-189. doi.org/10.1007/s13721-013-0029-7

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F.

(2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117(17), 9451-9457. doi.org/10.1073/pnas.1921046117

García, M. S. & Grijalva, C. Q. (2021). Anticancer Secondary Metabolites Found in Native Ecuadorian Plant Species *Uncaria tomentosa* DC. (Rubiaceae), *Croton lechleri* Müll. Arg. (Euphorbiaceae), and *Equisetum giganteum* L. (Equisetaceae). *Innovations in Biotechnology for a Sustainable Future*, 377-401. doi.org/10.1007/978-3-030-80108-3_18

Girlanda, M., Segreto, R., Cafasso, D., Liebel, H. T., Rodda, M., Ercole, E., ... & Perotto, S. (2011). Photosynthetic Mediterranean meadow orchids feature partial mycoheterotrophy and specific mycorrhizal associations. American Journal of Botany, 98(7), 1148-1163. doi.org/10.3732/ajb.1000486

Gluck-Thaler, E. & Slot, J. C. (2018). Specialized plant biochemistry drives gene clustering in fungi. *ISME*. 12, 1694-1705. doi.org/10.1038/s41396-018-0075-3

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075. doi.org/10.1093/bioinformatics/btt086

Higginbotham, S. J., Arnold, A. E., Ibanez, A., Spadafora, C., Coley, P. D., & Kusar, T. A. (2013). Bioactivity of Fungal Endophytes as a Function of Endophyte Taxonomy and the Taxonomy and Distribution of Their Host Plants. *PLoS ONE,* 8.(9) 10.1371/journal.pone.0073192

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., ... & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research*, *47*(D1), D309-D314.

Korf, I. (2004). Gene finding in novel genomes. *BMC bioinformatics*, 5(1), 1-9. doi.org/10.1186/1471-2105-5-59

Lukashin, A. V., & Borodovsky, M. (1998). GeneMark. hmm: new solutions for gene finding. *Nucleic acids research*, 26(4), 1107-1115. doi.org/10.1093/nar/26.4.1107

Maehara, S., Simanjuntak, P., Kitamura, C., Ohashi, K., & Shibuya, H. (2012). Bioproduction of Cinchona alkaloids by the endophytic fungus Diaporthe sp. associated with Cinchona ledgeriana. *Chemical and Pharmaceutical Bulletin*, 60(10), 1301-1304. doi.org/10.1248/cpb.c12-00545

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics, 20(16), 2878-2879. doi.org/10.1093/bioinformatics/bth315

Marin-Felix, Y., Miller, A. N., Cano-Lira, J. F., Guarro, J., García, D., Stadler, M., Huhndorf, S. M., & Stchigel, A. M. (2020). Re-Evaluation of the Order Sordariales: Delimitation of Lasiosphaeriaceae s. str., and Introduction of the New Families Diplogelasinosporaceae, Naviculisporaceae, and Schizotheciaceae. *Microorganisms*, 8.(9), 1430. doi:10.3390/microorganisms8091430

Monteiro, P. C. M., Alves, N. M., Queiroz, M. V., Pinho, D. B., Pereira, O. L., Souza, S. M. C., & Cardoso, P. G. (2017). Antimicrobial Activity of Endophytic Fungi From Coffee Plant. *Bioscience Journal,* 33.(2), 381-389.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution,* 32(1), 268-274. doi.org/10.1093/molbev/msu300

Oide, S., & Turgeon, B. G. (2020). Natural roles of nonribosomal peptide metabolites in fungi. *Mycoscience,* 61(3), 101-110. doi.org/10.1016/j.myc.2020.03.001

Palmer J. 2015. Funannotate. https://github.com/nextgenusfs/funannotate.

Prjibelski, Andrey, et al. "Using SPAdes de novo assembler." *Current protocols in bioinformatics* 70.1 (2020): e102. doi.org/10.1002/cpbi.102

Sachin, N., Manjunatha, B. L., Kumara, P. M., Ravikanth, G., Shweta, S., Suryanarayanan, T. S. Ganeshaiah, K. N., & Shaanker, R. U. (2013). Do endophytic fungi possess pathway genes for plant secondary metabolites? *Current Science,* 104.(2), 178-182.

Salazar-Cerezo, S., Martínez-Montiel, N., García-Sánchez, J., Pérez-y-Terrón, R., & Martínez-Contreras, R. D. (2018). Gibberellin biosynthesis and metabolism: A convergent route for plants, fungi and bacteria. *Microbiological research,* 208, 85-98. doi.org/10.1016/j.micres.2018.01.010

Sarrocco, S. (2016). Dung-inhabiting fungi: a potential reservoir of novel secondary metabolites

for the control of plant pathogens. *Pest Management Science*, 72, 643-652. 10.1002/ps.4206

Schulz, M., Knop, M., Muellenborn, C., & Steiner, U. (2013) Root-Associated Microorganisms Prevent Caffeine Accumulation in Shoots of *Salvia officinalis* L. *International Journal of Agriculture and Forestry*, 3.(4), 152-158. DOI: 10.5923/j.ijaf.20130304.05

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212. doi.org/10.1093/bioinformatics/btv351

Slot, J. C. (2017). Fungal Gene Cluster Diversity and Evolution. *Advances in Genetics*, 100, 141-178. doi.org/10.1016/bs.adgen.2017.09.005

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic acids research*, 32(suppl_2), W309-W312. doi.org/10.1093/nar/gkh379

Sugiyama, A., Sano, C. M., Yazaki, K., & Sano, H. (2016). Caffeine fostering of mycoparastic fungi against phytopathogens. *Plant Signaling & Behavior*. 11.(1). doi.org/10.1080/15592324.2015.1113362

Sukumar, P., Legue, V., Vayssieres, A., Martin, F., Tuskan, G. A., & Kalluri, U. C. (2013). Involvement of auxin pathways in modulating root architecture during beneficial plant–microorganism interactions. *Plant, cell & environment*, 36(5), 909-919. doi.org/10.1111/pce.12036

Tennakoon, D. S., Kuo, C. H., Maharachchikumbura, S. S., Thambugala, K. M., Gentekaki, E., Phillips, A. J., ... & Hyde, K. D. (2021). Taxonomic and phylogenetic contributions to Celtis formosana, Ficus ampelas, F. septica, Macaranga tanarius and Morus australis leaf litter inhabiting microfungi. *Fungal Diversity*, 1-215. 10.1007/s13225-021-00474-w

Valone, J. A., Case, M. E., & Giles, N. H. (1971). Constitutive mutants in a regulatory gene exerting positive control of quinic acid catabolism in Neurospora crassa. Proceedings of the National Academy of Sciences, 68(7), 1555-1559.

Vouffo, E. D., Douanla-Meli, C., Awantu, A. F., Lenta, B. N., Ngouela, S., Tsamo, E., & Lasstsch, H. (2021). New Metabolites From the Endophytic Fungus *Cecophora samala* Associated With *Mitragyna inermis*. *Natural Product Communication*, 16.(5), 1-6. 10. 1177/ 1934 578X 2110 13220

Wang, B., Liang, X., Gleason, M. L., Zhang, R., & Sun, G. (2017). Genome sequence of the ectophytic fungus Ramichloridium luteum reveals unique evolutionary adaptations to plant surface niche. *BMC genomics*, 18(1), 1-14. doi.org/10.1111/lam.13571

Wheeler, T. J., Clements, J., Eddy, S. R., Hubley, R., Jones, T. A., Jurka, J., ... & Finn, R. D. (2012). Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic acids research*, 41(D1), D70-D82. doi.org/10.1093/nar/gks1265

Wisecaver, J. H., & Rokas, A. (2015). Fungal metabolic gene clusters—caravans traveling across genomes and environments. *Frontiers in Microbiology*, 6, 161. doi.org/10.3389/fmicb.2015.00161

Wisecaver, J. H., Slot, J. C., & Rokas, A. (2014). The evolution of fungal metabolic pathways. *PLoS genetics*, 10(12), e1004816. doi.org/10.1371/journal.pgen.1004816

Yao , C. & Köller, W. (1993). Diversity of cutinases from plant pathogenic fungi: cloning and characterization of a cutinase gene from *Alternaria brassicicola*. *Physiological and Molecular Plant Pathology*, 44.(2), 81-92. doi.org/10.1016/S0885-5765(05)80103-1